PROTEINS: Structure, Function, and Bioinformatics 57:311–321 (2004)

Gibbs Adsorption Isotherm Combined with Monte Carlo Sampling to See Action of Cosolutes on Protein Folding

Daniel Harries* **and V. Adrian Parsegian**
Laboratory of Physical and Structural Biology, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland

ABSTRACT **Driven by conditions set by smaller solutes, proteins fold and unfold. Experimentally, these conditions are stated as intensive variables - pH and other chemical potentials - as though small solutes were infinite resources that come at an externally varied free energy cost. Computationally, the finite spaces of simulation allow only fixed numbers of these solutes. By combining the analytic Gibbs adsorption isotherm with the computational Monte Carlo sampling of polymer configurations, we have been able to overcome an inherent limitation of computer simulation. The idea is to compute analytically the free energy changes wrought by solutes on each particular configuration. Then numerical computation is needed only to sample the set of configurations as efficiently as when no bathing solute is present. For illustration, the procedure is applied to an idealized two-dimensional heteropolymer to yield lessons about the effect of cosolutes on protein stability. Proteins 2004;57:311–321.**

INTRODUCTION

Inside cells, proteins fold spontaneously into specific, biologically functional, native forms. The self-assembled structure is determined not only by amino-acid sequence but also by the cellular environment of other proteins, nucleic acids, lipids and smaller molecules. Malleable polymers, proteins respond differentially to various agents that affect the stability of different states or configurations.[1–5] *In vitro* measurements show that these cosolutes act on proteins through crowding, osmotic stress and other preferential interactions.[6–15] Some cosolutes, such as sugars, can further stabilize the native state of a protein.[3,9,11,16,17] Others, chemical denaturants, such as urea and guanidinium, destabilize the native state to favor unfolded (denatured) conformations. [10,18–22] Some agents even destabilize particular protein conformations but stabilize new, well ordered, secondary structures (such as helixes).[23]

Because of the large number of solute and cosolute molecules that must be considered for reliable thermodynamic averages, it can be difficult to model proteins under the constraints of fixed chemical potentials.[24,25] Several methods treat the solvent as a mean field acting to affect protein energy.[26–29] Instead, we develop here an efficient way to evaluate the equilibrium ensemble of protein states in a mixed (water-cosolute) solution without making mean-field assumptions. Combining thermodynamic averaging with Monte Carlo (MC) simulation,[30] we use the analytic expressions for the adsorption isotherm[31] to evaluate the free energy change from cosolute–protein interaction at each instantaneous protein configuration. The strategy is somewhat similar to that used in recent years to study the native structure of proteins at fixed pH, taking into account in the simulations the charged state of side chains according to the proton chemical potential in solution.[32–35] However, we use a MC procedure to sample the equilibrium ensemble of states under different chemical potentials, to follow the process of protein folding or unfolding in response to its environment.

Specifically, by evaluating the adsorption free energy of cosolutes acting on each configuration, we are able to keep the chemical potential of the bathing solution fixed while testing each protein conformation generated by a MC move. As usual, the weight of each configuration is evaluated according to its Boltzmann probability in free energy, but it is now based on the sum of inter-protein energy and cosolute adsorption free energy. In this way, we preclude the need for MC sampling of all the different cosolute–protein configurations.

We develop and illustrate this computational strategy by considering the example of a heteropolymer on a two-dimensional (2D) lattice. In the spirit of the so-called 'HP model',[36,37] the protein is represented as a self-folding polymer, composed of hydrophobic and hydrophilic (or polar) monomers. In a set of MC moves, such proteins can self-assemble into one or few compact, energetically distinct, native structures. In the model, cosolutes are small, each assumed to occupy a single lattice site. Each cosolute has preset adsorption (free) energies with the two types of protein monomers, reflecting the overall tendency of cosolute to associate with a monomer as compared to its tendency to remain solvated in the bathing solution.

Our goal in this illustration is to reveal the link between the preferential interaction of cosolute with protein and the ensuing change in protein stability and conformation once cosolute and protein interact. First, we present a simple model that captures essential features of protein

*Correspondence to: Daniel Harries, Laboratory of Physical and Structural Biology, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892-0924. E-mail: harries@helix.nih.gov

Received 15 December 2003; Revised 16 March 2004; Accepted 25 March 2004

Published online 11 June 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20182

denaturation or stabilization by cosolutes. Second, as expected from the excess/deficit reasoning of the Gibbs adsorption isotherm, changes in protein conformation and stability are demonstrably reflected in the directly counted number of associated cosolutes. Finally, by comparing protein conformations - both native and denatured - and the corresponding levels of associated solute in the presence of different cosolutes, it becomes clear how protein structure is determined by the extent and characteristics of protein–cosolute association.

METHOD

For concreteness, we develop the method in terms of a specific model, a protein polymer of N monomer amino acids allowed to move on a 2D square lattice of M cells. The lattice contains only one chain. With the consequent absence of inter-chain interactions, the system is in effect infinitely dilute with respect to protein.

In addition to the single protein chain, there are n_s small cosolute molecules, with each cosolute allowed to occupy a single lattice site. The remaining n_w lattice sites are (implicitly) associated with solvent (e.g., water) so that $M = N + n_s + n_w$.

We assign to each monomer either hydrophobic (H) or polar (P) propensity. When they occupy nearest-neighbor sites on the lattice, two non-consecutive (non-bonded) monomers of type H interact with a favorable energy $\epsilon_{hh} \leq 0$. In this HP model, the cosolutes interact with the chain's H- and P-type monomers through nearest-neighbor interactions of energies ϵ_{sh} and ϵ_{sp}, respectively. We assume ϵ values are independent of neighboring occupancy and set all other possible interaction energies to zero.

We work in the "semi-grand-canonical ensemble".[38] That is, each cell in the lattice is permeable to cosolute and solvent molecules. The number of cosolute molecules in the simulation box is determined by coupling to a bath where the cosolute chemical potential is set at μ^s for a concentration $c_s = n_{s,b}/M_b$ in a protein-free bath lattice of M_b cells containing on average $n_{s,b}$ cosolutes. Assuming that the solution in the bath lattice is ideal, $\mu^s = k_B T \ln[c_s/(1 - c_s)]$ (see e.g. ref. 39).

The partition function $\mathcal{Z}$ for the protein is expressed in terms of the free energies $E_i + g_i^s$ of individual conformations i. Here, E_i is the sum of monomer–monomer interaction energies, and g_i^s accounts for all other protein–cosolute interactions. The sum

$$\mathcal{Z} = \sum_{i=1}^{\Omega} e^{-\beta(E_i + g_i^s)} \tag{1}$$

extends over all possible polymer conformations; $\beta \equiv 1/k_B T$, T is the absolute temperature, and k_B is Boltzmann's constant.

Provided N is not too large, it is often possible to enumerate all protein configurations explicitly.[37] Similarly, even all cosolute–protein configurations might be enumerated for each configuration so as to evaluate the partition function exactly. Here, in order to explore systems with larger N values, we use MC moves to generate an ensemble of protein conformations.

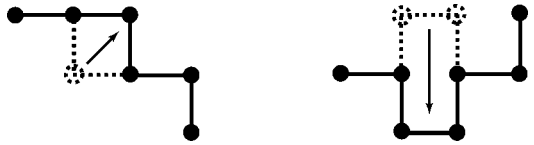


Fig. 1. Types of MC attempts. At positions away from the chain termini, one-segment or two-adjacent-segment flips are tried at randomly chosen positions. At either terminus, simultaneous moves of the three last segments are tested.

Each trial MC move is made by pivoting one or two monomers chosen at random while two flanking segments are kept fixed. Such moves incur shifts of one or two or three bonds (see Fig. 1). At the two chain terminal regions, moves are made at random to the three terminal monomers in attempts to change their conformation. Trial polymer configurations that result in overlapping monomers are rejected; for other moves, the change in internal energy (between protein monomers) is evaluated.

Rather than using insertion and removal moves of cosolute particles into and out of the system to maintain fixed chemical potential, we here use a different approach. For any protein configuration i generated using the MC procedure, we evaluate the (free) energy g_i^s associated with the cosolutes–polymer interaction. We do this by evaluating the (grand canonical) partition function for each adsorption site j that neighbors one or more of the protein monomers (see e.g. ref. 39):

$$\xi_{ij} = 1 + e^{-\beta(\epsilon_j - \mu^s)} \tag{2}$$

For a specific polymer configuration i, the cosolute-related free energy is therefore

$$g_i^s = -k_B T \ln \prod_{j=1}^{2N+2} \xi_{ij} \tag{3}$$

In eq. (3), the product extends over the maximum number of interaction sites that may become available to the cosolute, $2N + 2$. Note that in most conformations not all sites will be available for binding: typically some will be occluded from a cosolute's reach. For those sites, we assume $\epsilon_j = 0$, $\xi_{ij} = 1 + e^{\beta\mu^s}$ (as we do for other sites in the bathing solution away from the polymer). It is important to take all $2N + 2$ sites into account so as to keep a well-defined reference free energy, taken here to be the fully extended chain with $2N + 2$ sites exposed to cosolute.

The transition to a new trial configuration in the MC procedure is determined by the usual Metropolis procedure[30] with the Boltzmann weight of $\exp[-\beta(g_i^s + E_i)]$. We thus trace over the cosolute degree of freedom for each polymer configuration while a new trial configuration is generated during the MC run. In this sense, we evaluate (free) energies quickly, avoiding the need to evaluate them for all possible protein conformations with all cosolute complexions.

After an initial equilibration period, typically $\approx 10^6$ MC moves (each move, on average, including one attempted trial for each protein segment), conformations are recorded for at least another 10^7 moves. In order to ensure two configurations sampled are independent, data are collected only every 10^2 MC steps. By following the ensemble of probable states once the simulation has reached equilibrium, the number of associated cosolutes with a conformation i can be evaluated. We do this by summing the probability for each (potential) binding site j on the protein's surface to be occupied by cosolute,

$$n_i = \sum_{j=1}^{2N+2} \frac{e^{-\beta(\epsilon_j - \mu^s)}}{1 + e^{-\beta(\epsilon_j - \mu^s)}} \tag{4}$$

In order to obtain the excess of bound cosolute $n_i^{es} = n_i - n_b$ from the evaluated n_i, we must account for the background concentration of bathing cosolute in the absence of protein $n_b = (2N + 2)e^{\beta\mu^s}/(1 + e^{\beta\mu^s})$. Note here that, out of the $2N + 2$ possible sites, the sites unavailable for cosolute in many protein conformations are subtracted out in this way.

We compared this simulation approach with another, where explicit cosolutes were modeled using insertion and annihilation trial moves to keep the chemical potential fixed. We found that both methods gave identical results within statistical error (see below). However, the time needed to run the traditional simulation was much longer. This rapidity is possible in the simulation method employed here because the run time becomes independent of μ^s. The number of terms in the partition function depends only on N, not on c_s.

Equilibrium Constant and Free Energy

It has been shown that within the HP model one can design many chain sequences that, for energetic reasons, reveal a single most stable conformation at thermodynamic equilibrium. If such a conformation exists, we define it as the native state ($\mathcal{N}$). All other conformations will be defined as part of the denatured state ($\mathcal{D}$). In these terms, the equilibrium constant for denaturation, $\mathcal{N} \rightleftharpoons \mathcal{D}$, is defined as the equilibrium ratio between ensemble probabilities for the denatured state, $P_{\mathcal{D}}$, and the native state, $P_{\mathcal{N}}$,

$$K_{\mathcal{N}\mathcal{D}} = \frac{P_{\mathcal{D}}}{P_{\mathcal{N}}} = \frac{\sum_{i=2}^{\Omega} P_i}{P_1} \tag{5}$$

In eq. (5) we have set $i = 1$ for the native state and $i = 2, \ldots, \Omega$ for all other (denatured) conformations. The standard free energy for the reaction is then

$$\Delta G = -k_B T \ln K_{\mathcal{N}\mathcal{D}} \tag{6}$$

We focus specifically on $\delta\Delta G(c_s) \equiv \Delta G(c_s) - \Delta G(c_s = 0)$, the difference in ΔG due to cosolute addition.

The change in the free energy for a certain protein conformation upon immersion in a cosolute bathing solution at a chemical potential μ^s is related to the excess (average) number of cosolutes in the vicinity of that configuration, N^{es}:

$$dG_i^{\text{immersion}} = -N^{es} d\mu^s \tag{7}$$

The relative stability of the (single) native configuration and all other denatured configurations is then related to the difference in the number of excess cosolutes in the two states,

$$\frac{d\delta\Delta G}{d\mu^s} = -\Delta N^{es} \tag{8}$$

This is the well-known Wyman linkage relation.[40] For the collection of denatured configurations, we in fact consider the ensemble average of associated cosolutes in that state n_i^{es}; each configuration contributes to the count of excess cosolutes in proportion to its equilibrium probability $N^{es} = \langle n_i^{es} \rangle_T$.

A Measure of Nativeness

There is no unique way to define the extent to which a given configuration differs from the native configuration. Here, we find it most instructive to speak of the "native-state overlap" Q.[41] This measure allows us to assign probabilities for configurations containing a certain number of native inter-protein contacts. We found that, on the nativeness scale, this measure closely groups configurations that are not only structurally similar but also have a similar ensemble probability at equilibrium.

We define the overlap by first introducing Δ^m, the contact matrix for a certain configuration m, whose element Δ_{ij}^m is 1 if monomers i and j are non-consecutive monomers in contact (i.e., nonbonded nearest neighbors on the lattice), and 0 otherwise. The contact matrix is symmetric, and all its diagonal, sub- and superdiagonal elements are zero. We denote the contact matrix for the native state $\mathcal{N}$ by its elements $\Delta_{ij}^{\mathcal{N}}$. For any configuration m, the native-state overlap is formally defined here as

$$Q_m = \sum_{\langle i,j \rangle} \Delta_{ij}^{\mathcal{N}} \Delta_{ij}^m \tag{9}$$

with the sum extending over all i, j pairs once. Simply stated, for a native state overlap score, a monomer–monomer contact is counted if it appears also in the native configuration. We will present results in terms of $P(Q)$, the equilibrium ensemble probability for any configuration to score a value Q.

Protein

As with real proteins, different model proteins are expected to follow different folding/unfolding pathways and to display intermediates that reflect the sequence and native structure of the protein. Here, to follow specifically the role of cosolute, we limit consideration to one protein sequence exposed to different cosolutes and concentrations.

We choose a protein chain, Figure 2, extensively studied for its distinct folding and intermediates.[42–44] Composed of 20 monomers ($N = 20$), 12 of type P and 8 of type H, it

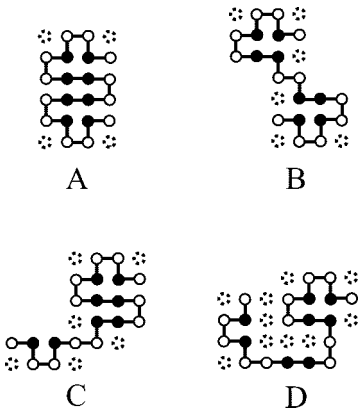


Fig. 2. Compactness seen in native (A) and several typical denatured (B–D) conformations. Polar monomers (P) are indicated by open circles and hydrophobic (H) by filled circles. The broken circles show sites of possible cosolute binding to two or more monomers. The native conformation (A) enjoys a maximum of eight HH contacts and a native-state overlap score of $Q = 11$. Denatured conformation B has six HH contacts and $Q = 6$; conformation C, 6 HH contacts with $Q = 7$; conformation D, 4 contacts with $Q = 4$ [computed using eq. (9)].

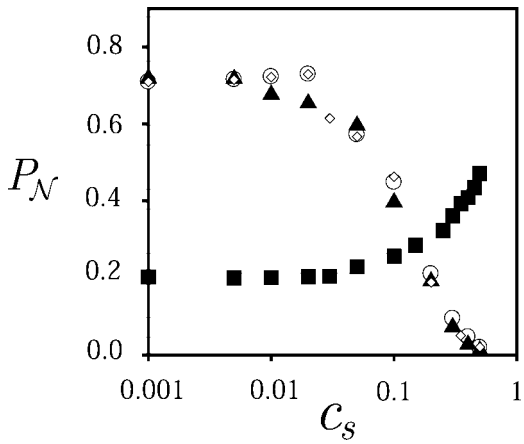


Fig. 3. Probability of the native configuration as a function of cosolute concentration. Different sets of interaction parameters reveal different forms of response. Squares (turning up) show stabilization of weakly folded native form by exclusion of repelled cosolute, (HH segment contact energy $\epsilon_{hh} = -3k_BT$, cosolute to segments H and P contact energies, $\epsilon_{sh} = \epsilon_{sp} = 1k_BT$). Circles and triangles (turning down) show compact native form undone by absorptive cosolute. For triangles $\epsilon_{hh} = -4k_BT$, $\epsilon_{sh} = 0$, $\epsilon_{sp} = -2k_BT$. For circles $\epsilon_{hh} = -4k_BT$ and $\epsilon_{sh} = \epsilon_{sp} = -1k_BT$. For comparison, shown in diamonds, results from simulations using the usual MC grand canonical ensemble method, $\gtrsim 10^2$ times slower, involving explicit cosolutes with the same contact energies as used for circles.

spontaneously folds to the native structure [Fig. 2(a)] in a few MC steps. In the native state, all H monomers are surrounded by P monomers, so that no H monomer is exposed to the bathing solution. The native structure has been described as a 2D analog of a β-sheet and has been shown in MC simulations to fold through several well-defined intermediates and pathways.[42–44] The energy of the native conformation is at least $2\epsilon_{hh}$ lower than that of any other unfolded conformation. Consequently, for $\epsilon_{hh} \lesssim -3k_BT$, this is the most stable configuration so that $\Delta G < 0$. However, this does not imply that the system contains only two well-defined states. In fact, we show in the following sections that other stable collections of configurations, intermediates, may become populated under different conditions.

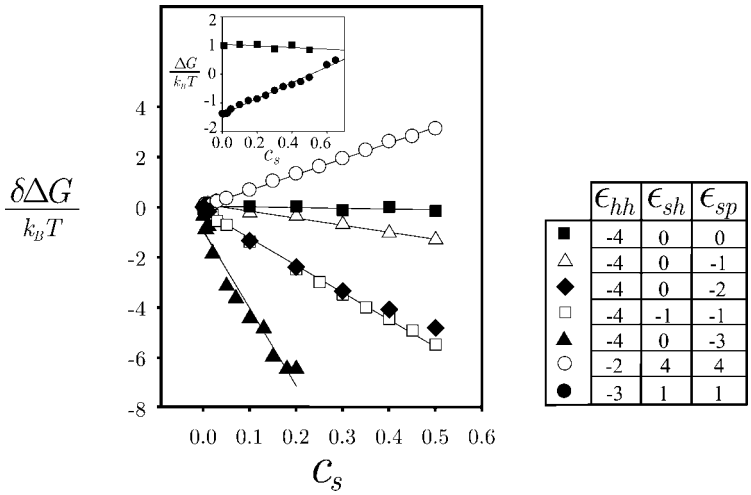
This model protein has also been studied in the context of chemical denaturation; the denaturant effect was introduced through a change in the HH interaction parameter ϵ_{hh}. With no other interaction in the system, this change in ϵ_{hh} corresponds to a change in temperature; lowering ϵ_{hh} is equivalent to raising T. Working through ϵ_{hh} may also be considered a mean field approximation to the denaturation with a single field (temperature) acting to denature the protein. In contrast, in this study, we vary cosolute chemical potentials as well as cosolute–protein interaction parameters, in this way locally affecting individual monomers to different extents. With many more, explicitly-evaluated, local fields acting on the protein, we expect a richer denaturing behavior. We can also explicitly account for numbers of associated cosolutes and follow the relationship between stability and numbers adsorbed using Wyman linkage [eq. (8)]. Where instructive, we will compare thermal to cosolute denaturation.

RESULTS

Protein Stability

For several types of cosolutes, Figure 3 shows the probability $P_{\mathcal{N}}$ of finding the native state in the ensemble of accessible configurations. As the chemical potential of cosolute in the bath is varied, the probability (hence stability) of the native state changes, reflecting the protein's susceptibility to solution composition. Added cosolute that interacts unfavorably or is excluded from the protein's neighborhood ($\epsilon_{sh} = \epsilon_{sp} = 1k_BT$, squares in Fig. 3) tends to stabilize the compact native configuration. The protein minimizes the surface exposed to interactions. In contrast, agents that act favorably toward one or two types of monomers tend to denature the protein and to favor non-native conformations. Figure 3 also shows the action of two denaturing agents that are very similar in their denaturing ability, although their interaction parameters differ quite substantially. While one ($\epsilon_{sh} = 0$, $\epsilon_{sp} = -2k_BT$, triangles in Fig. 3) interacts with P type monomers only, the other ($\epsilon_{sh} = \epsilon_{sp} = -1k_BT$, circles in Fig. 3) interacts with both types of monomers to the same extent. In spite of similar effects on protein stability, the ensuing protein denatured states differ (shown below).

Finally, Figure 3 shows that the simulation method we present gives results identical to those that are obtained using the common grand-canonical ensemble MC requiring periodic attempts to delete and insert cosolute into the simulation box. In Figure 3, diamonds correspond to simulations using such explicit cosolutes, while circles show results using our implicit coslute method for the same interaction parameters. The results of both are identical within statistical error. However, run time for high cosolute density ($c_s \gtrsim 0.1$) was at least 100 times longer using explicit cosolute. In the following sections, we demonstrate thermodynamic equalities in the system - most importantly the Gibbs adsorption isotherm - showing that equilibrium is fully attained in simulations using implicit cosolute.



	ϵ_{hh}	ϵ_{sh}	ϵ_{sp}
■	-4	0	0
△	-4	0	-1
◆	-4	0	-2
□	-4	-1	-1
▲	-4	0	-3
○	-2	4	4
●	-3	1	1

Fig. 4. Differences in protein stability, $\delta\Delta G$ caused by addition of several different cosolutes to proteins with the same sequence. Inset: Stabilization by the action of cosolutes. Change in denaturation free energy ΔG vs. added cosolute for two proteins of the same sequence but different contact energies. Note that in the inset $\Delta G(c_s)$ from Eq. 6 is itself plotted, not the differences between $\Delta G(c_s)$ and $\Delta G(c_s = 0)$. Interaction parameters for the different sets are tabulated in units of $k_B T$.

For all cosolutes, the dependence of the $\mathcal{N}$ to $\mathcal{D}$ transition on cosolute concentration is rather sharp, though not as sharp as in many real proteins. We may attribute this difference to the coarse grained, 2D nature of our model. Yet some important features of protein denaturing are reproduced faithfully. One of these is the dependence of protein denaturing free energy $\delta\Delta G = \Delta G(c_s) - \Delta G(c_s = 0)$ on cosolute concentration, shown in Figure 4.

We find that, for a large number of cosolute-interaction parameters, protein stability is nearly linear in cosolute concentration. This is true both for denaturing and for stabilizing agents. Because it is possible within the simulation to evaluate the equilibrium constant directly for the case of no added cosolute (a complex feat in real experiments), we can conclude that within our model this linearity is maintained from low to high cosolute concentrations. An exception to this is the cosolute that interacts strongly with the P type monomers only ($\epsilon_{sh} = 0$, $\epsilon_{sp} = -3k_B T$, Fig. 4 full triangles), for which deviation from linearity greater than the statistical error is observed. We can expect such deviations for strongly associating cosolutes. In the extreme case of very strong adsorption, this non-linearity indicates the finite number of accessible, cosolute adsorption sites and depends strongly on the details of protein structure.

The strong influence of stabilizing cosolutes on the protein can be easily appreciated by considering the changes in ΔG itself (rather than $\delta\Delta G$). The inset in Figure 4 follows the change in protein stability for two proteins with the same monomer sequence but different internal interaction parameters ϵ_{hh}. For one, $\epsilon_{hh} = -3k_B T$, making the native state unfavorable, so that in the absence of cosolute $P_{\mathcal{N}} \approx 0.2$ (circles in Fig. 4 inset). As excluded cosolutes ($\epsilon_{sh} = \epsilon_{sp} = +1k_B T$) are added, the probability of the compact native state increases (ΔG for denaturation becomes more positive). For comparison, we show results for the same protein sequence stabilized by the inter-monomer interactions $\epsilon_{hh} = -4k_B T$. As expected, the native state is the stable state ($P_{\mathcal{N}} \approx 0.75$), remaining unchanged (within statistical error) in the presence of cosolutes with which it has zero interaction energy (Fig. 4 inset, filled squares).

At higher cosolute concentrations, the poorly folded protein becomes better folded, until at $c_s \approx 0.6$ it (almost) regains the stability of the reference protein. By its unfavorable interaction with the protein, the cosolute works to stabilize the compact native state. Forming a cosolute-free zone around the protein, the system minimizes energetically unfavorable cosolute–protein interactions. However, in so doing, the system incurs another price associated with the loss of mixing entropy from the cosolute excluding volume. As the cosolute concentration in the bath is elevated, this price becomes larger; the system tends to minimize the penalty by compacting the protein, thus minimizing the excluding volume (solvent, empty lattice sites) the protein produces. In this sense this is a solvent (water) effect (rather than a cosolute effect), though the solvent is only implicitly taken into account here.[7]

Cosolute Adsorption

By studying the way in which different cosolutes adsorb onto the protein under different conditions, we learn how adsorption confers stability. For two cosolutes, Figure 5 follows the total number of cosolutes adsorbed to the protein (squares) as well as the number of exposed protein sites available for interaction (circles) as cosolute is added. Though the effect of these two denaturing agents on protein stability is similar (compare open squares and diamonds, Figure 4), their adsorption behavior is different, as compared below.

In Figure 5, filled symbols correspond to weak cosolute interaction with both P and H type monomers with the same energy, $\epsilon_{sh} = \epsilon_{sp} = -1k_B T$. In this case, we observe a

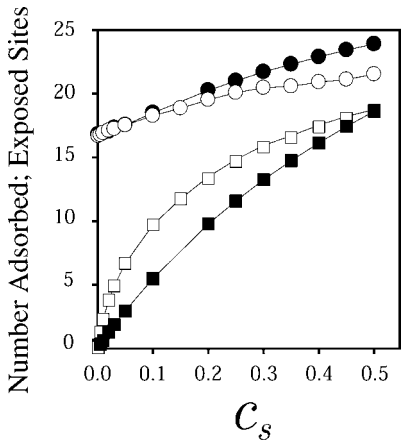


Fig. 5. Cosolute action on folding seen through adsorption isotherms. Squares show ensemble-average total numbers of cosolute adsorbed to all protein conformations; circles show number of exposed protein sites. For filled symbols, $\epsilon_{sh} = \epsilon_{sp} = -1k_BT$; for open symbols, $\epsilon_{sh} = 0$ and $\epsilon_{sp} = -2k_BT$. For both open and full $\epsilon_{hh} = -4k_BT$.

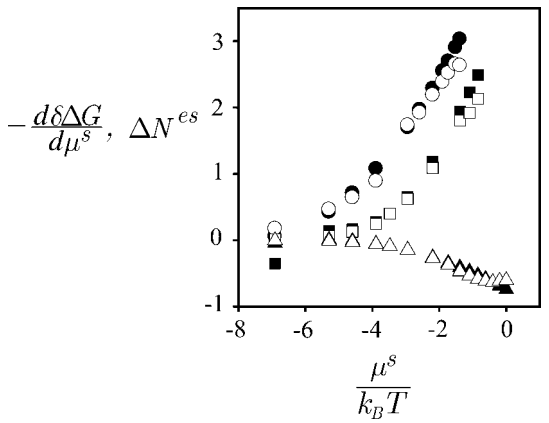


Fig. 6. Evidence of the adsorption isotherm: comparison of the changes in the number of excess cosolutes associated with the protein, ΔN^{es} (open symbols) and change in protein stability with the change in cosolute chemical potential, $d\delta\Delta G/d\mu^s$ (full symbols), as a function of cosolute chemical potential. Circles correspond to $\epsilon_{hh} = -4k_BT$, $\epsilon_{sh} = 0$ and $\epsilon_{sp} = -3k_BT$. Squares correspond to $\epsilon_{hh} = -4k_BT$ and $\epsilon_{sh} = \epsilon_{sp} = -1k_BT$. Triangles correspond to $\epsilon_{hh} = -3k_BT$ and $\epsilon_{sh} = \epsilon_{sp} = 2k_BT$.

gradual opening of new adsorption sites with increasing cosolute concentration. This reflects the tendency of the protein to make more sites available for favorable interaction with cosolutes. In concert, the numbers of cosolutes adsorbed also grows gradually. The curves do not seem to reach saturation at all studied concentrations, reflecting the fact that the protein is still undergoing unfolding. (Recall that the maximum number of sites on the protein is $2N + 2 = 42$.)

In contrast, $\epsilon_{sh} = 0$ and $\epsilon_{sp} = -2k_BT$ (Fig. 5, open symbols) show a Langmuir-like cosolute adsorption isotherm. This saturating adsorption reflects the nature of the protein's unfolding. Rather than continuously unraveling, the protein follows more closely a two state process; it reaches ~22 available sites for $c_s = 0.2$ (Fig. 5, circles) and remains almost unchanged after that. This number of sites in turn becomes the limit of cosolute adsorption.

The difference in the action of the two cosolutes considered is related to their interaction with the protein and the ensuing stabilization of different protein configuration ensembles (states or intermediates). The most probable configurations found in the presence of cosolute is determined by the balance of intermonomer and cosolute–monomer interactions. When cosolute tends to associate favorably with one monomer type but not with the other, the system is frustrated. Rather than unfolding completely to allow both maximal contact with cosolute and chain configurational entropy, the protein will only unfold to the extent that adsorbed cosolute to one monomer type can stabilize it at the expense of lost favorable intermonomer interactions.

In fact, we show in the next section that the action of different cosolutes can promote structurally different protein intermediates and denatured states, even when their stability relative to the native state is comparable. As a result, a final denatured state may present a different number of cosolute adsorbing sites in different cosolutes.

From Wyman linkage, eq. (8), we expect the number of cosolutes associated with the native versus all denatured configurations to reflect protein stability. In thermodynamic equilibrium, ΔN^{es} is equal to the change in protein stability at a certain cosolute chemical potential. Indeed, in the simulation we can monitor these two thermodynamic variables: numbers and free energies, in an independent manner. First, we follow ΔN^{es}, the difference in numbers of associated (included or excluded) cosolutes close to the protein (nearest neighbor) in excess with respect to their numbers in the bathing solution. We do so by taking ensemble averages, as described above [see eq. (4)]. Second, we follow protein stability extracted from the equilibrium constant $K_{\mathcal{N}\mathcal{D}}$ [eq. (6)]. By fitting the set of evaluated $\delta\Delta G$ for different μ_s to a 4th (or higher) order polynomial, we can access the derivative $d\delta\Delta G/d\mu^s$. We compare this derivative to ΔN^{es} for several different cosolutes in Figure 6.

For all cases considered, we find that numbers and changes in stability are identical within statistical error of the simulation. This is true for both included and excluded cosolutes. Note that the relative error between the two observables becomes slightly larger for $\mu^s \gtrsim -1k_BT$. One possible explanation is that it is harder to obtain stable averages at such high cosolute concentrations, which tend to slow the transitions between the many configurational states. Another possible cause is the limited accuracy of the fit of $\delta\Delta G$ from which the derivatives are obtained.

As confirmed, changes in stability are directly related to differences in numbers adsorbed in the two (denatured and native) states. How then do different protein–cosolute interactions affect cosolutes in their stabilizing/destabilizing properties? One aspect is the extent to which cosolutes will adsorb/associate with the protein. This is directly related to the apparent interaction parameters for the cosolutes, and leads to the observation of the Wyman linkage discussed in this section; the more cosolutes bind to a particular state, the more that state is stabilized. The other aspect is more subtle and is related to the ability of different cosolutes to stabilize different denatured interme-

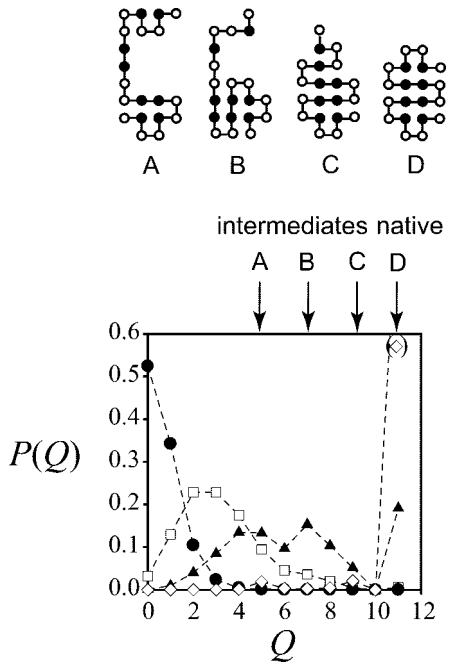


Fig. 7. Lower temperature and stronger H-H attraction compacts conformations. Probability of finding conformations with a native overlap Q [eq. (9)] for several temperatures (and no added cosolute), defined through the interaction parameter ϵ_{hh}. For diamonds, $\epsilon_{hh} = -6k_BT$, for triangles $\epsilon_{hh} = -3k_BT$, for squares $\epsilon_{hh} = -2k_BT$ and for circles $\epsilon_{hh} = 0$. Also shown are the native ($Q = 11$) and three typical denatured conformation corresponding to $Q = 9$, 7 and 5, as indicated by arrows. These configurations correspond to common structures in refs. 42 and 43.

diate states, depending on their mode of interaction with the protein, as discussed in the next section.

Protein Configurational Changes

Temperature Effects

Previous studies of the same model protein sequence used here not only followed the denaturation and folding but also showed the probable conformations for different folding conditions set by the interaction parameter ϵ_{hh}.[42–44] Figure 7 shows the probability of finding, at equilibrium, conformations with Q native contacts at different temperatures (characterized by ϵ_{hh}). Where no cosolute is present, the results closely follow previous studies of this protein. Though the method for classifying configurations is slightly different in our study, we find the predominance of the same types of conformations (for $\epsilon_{hh} = -4k_BT$). In fact, the same conformations appear in both studies as the leading probable denatured configurations (Fig. 7).

When only interaction parameter ϵ_{hh} acts, we can expect the internal (HH) protein to determine the stability of a certain configuration. This is indeed what we find when we observe the number of states that are most probable at different temperatures. As the temperature is elevated (ϵ_{hh} lowered), configurations with a smaller value of native contacts Q become more abundant. Moreover, a major determining factor in what the probable configurations will be (aside from the configurational degeneracy of a certain state) is the number of HH bonds in that configurations. Higher temperature configurations with a smaller number of HH bonds become more probable. In Figure 7, the most probable denatured configuration shifts from $Q = 9$ to $Q = 7$, $Q = 5$ and so on. Concomitantly, the number of intact HH bonds moves from 7 (only one less than in the native state) to 5 to 4. We conclude then that the number of HH bonds plays a pivotal role in determining the structure of the intermediate states of the protein under thermal denaturation.

Cosolute Stabilizing Effect Through Osmotic Stress

When cosolute is highly excluded, it is useful to consider the solvent chemical potential, rather than that of cosolute.[7] Using the Gibbs–Duhem relation, $n_s d\mu^s + n_w d\mu^w = 0$, we can rewrite eq. (8) to relate protein stability to the differences in numbers of cosolute excluding solvent between $\mathcal{D}$ and $\mathcal{N}$ states $\Delta N^{es} d\mu^s + \Delta N^{ew} d\mu^w = 0$,

$$\frac{d\delta\Delta G}{d\mu^w} = -\Delta N^{ew} \tag{10}$$

For low c_s, van't Hoff's law gives $d\mu^w = -k_BT dc_s$. We follow a highly excluded cosolute; shown in Figure 4 by empty circles. Here, linearity implies an osmotic (rather than cosolute binding) effect, enabling us to speak of a constant number of cosolute excluding (solvent) sites. First, from Figure 4, we find a loss of $\Delta N^{ew} = \beta(d\delta\Delta G/dc_s) \approx 7$ cosolute excluding solvent sites in the folding process. Independently, we find by counting directly in the simulation that the average number of sites neighboring the protein chain in the $\mathcal{D}$ state, when no cosolute is present, is ≈ 23.6. In the $\mathcal{N}$ state, only 16 sites are available for interaction with cosolute [see Fig. 2(a)]. The difference in available sites, ≈ 7.6, is very close to the 7 sites found from eq. (10).

We find that the stability of denatured configuration in an excluded cosolute corresponds simply to the number of exposed sites they display for unfavorable cosolute interaction. The main tendency to fold is then a matter of minimizing unfavorable contacts. Because the folded protein has the most favorable ratio of exposed to buried sites, it is also the one that becomes yet more stable.

Cosolute Denaturing Effect

When can we expect denaturation by cosolutes to stabilize intermediates that are similar to those obtained through thermal denaturation? Figure 8(a) shows the effect of one such cosolute, characterized by equal and weak interaction with both H and P monomers ($\epsilon_{sh} = \epsilon_{sp} = -1k_BT$). The cosolute effect is homogeneous over the whole surface of the protein and is analogous to that of a good versus bad solvent in theories of polymer swelling and protein denaturation;[46–47] the cosolute's action can be described adequately through one parameter (such as the chemical potential of the cosolute). Comparing Figure 7 and Figure 8, we find that, in terms of the most probable Q values, the stable intermediates in cosolute denaturation are quite similar to those formed in thermal denaturation. However, some differences can still be observed. In thermal denaturation, the internal HH energy is key. For

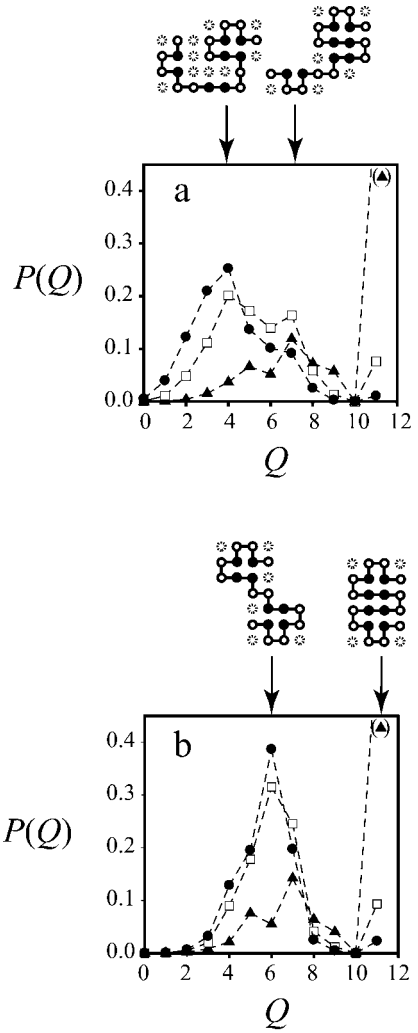


Fig. 8. Through protein-cosolute interactions, different cosolutes determine the protein's stable intermediates. Probability of finding conformations with a native overlap Q for several cosolute interaction parameters and concentrations. For both panels $\epsilon_{hh} = -4$. In (a) $\epsilon_{sh} = \epsilon_{sp} = -1k_BT$. In (b) $\epsilon_{sh} = 0$ and $\epsilon_{sp} = -2k_BT$. The different data sets in (a) and (b) correspond to: $c_s = 0.01$ triangles, $c_s = 0.3$ squares and $c_s = 0.5$ circles. The protein conformations correspond to the native structure and to typical configurations of the stable intermediate states.

cosolute denaturation, the ability of a certain configuration to enjoy two or more contacts with a single cosolute molecule (bridging interaction) is an additional important factor. Such sites add stability through favorable interaction energies with the cosolutes. Therefore, the system favors configurations that possess cavities, pockets or loops that enable cosolutes to interact with more than one monomer at a time [see Fig. 2(c and d)]. This aspect is hard to capture in mean-field theories. It corresponds to local structures (formation of loops and bridging interaction) and their ability to interact with the cosolute, depending on the details of the protein sequence.

When cosolute–protein interaction is stronger, or when they are very different for H and P type monomers, other effects on conformation stability can be seen. Figure 8(b) follows the probability of intermediates for $\epsilon_{sh} = 0$ and $\epsilon_{sp} = -2k_BT$. One striking feature is that a single Q value ($Q = 6$) becomes the most probable configuration. The ensemble of denatured states follows closely a two-conformation transition between those depicted in the figure. Other, more random configurations never become more favorable. This highly preferred configuration is a direct consequence of the specific cosolute–protein interaction. The system tends to expose P type monomers (favorably interacting with the cosolute) while burying H monomers (which interact with each other but not with the cosolute). The optimum is achieved with configuration B in Figure 2. This configuration maintains six of the eight possible HH interaction, while exposing many more P monomers for interaction, including some favorable cavities, enabling two P monomers to interact with a single cosolute.

The stabilizing effect of the cosolute on this particular conformation is also the reason for the fast saturation in cosolute adsorption and limited unfolding observed in Figure 5. The protein denatures only to the extent that the one stable conformation is unfolded, exposing the limited number of adsorption sites on that stable intermediate. Thus, structure is maintained in the denatured state, differing markedly from that found in the presence of the other cosolute shown in Figure 8(a).

Discussion

In recent years, it has been realized that understanding the conditions under which proteins fold is key to understanding their structure in solution. If we are to model proteins - whether natural or synthetically engineered - we must be able to model the effect of environmental conditions. Recall how cosolutes affect the stability of proteins: they do so through favorable or unfavorable preferential interactions; they tend to accumulate, solvate better or adsorb in the vicinity of the protein or may otherwise become depleted from the protein's surface. A manifestation of these interactions can be evaluated experimentally (for example by changes in solution vapor pressure due to protein addition) and are usually reported in terms of so-called 'preferential interaction coefficients'.[5,11,12] In turn, these effective potentials could be fed back into simulations of the type described here, to give predictions of structure and stability of proteins. Another important source for interaction parameters are atomic detail simulations of solvated peptides or small proteins and other macromolecules.[23,48–50]

Using a novel method for evaluating cosolute–protein complexions and their probability in equilibrium, we can follow the effects of cosolutes on proteins. Our technique is quite general and can be used whenever there is a way to evaluate the cosolute partition function for each particular protein conformation. Most models that take into account explicit cosolute will scale in computer run-time with the number of cosolute particles in the system. Here, because evaluations are made only according to the number of available exposed interaction sites, our algorithm speed is independent of the total number of particles. The result is

a reduction in run times, typically at least 100-fold for (physiologically) reasonable cosolute concentrations.

What can we expect to learn from simulations that account for the effect of cosolute? Here, we demonstrated some attainable properties in a lattice model system. In spite of their simplicity and inability to reproduce all the details of the experimentally observed folding process, lattice models have been used extensively to obtain some important insights on protein folding pathways and stability.[2,36,42,44,51–55] They have also been used to explore the binding of small agents to the protein, causing ligand-induced denaturation and other types of binding behaviors.[37] The model oversimplifies the interaction of cosolutes with proteins. For example, it is probably hydrated cosolute that associates with the protein's surface (or domain) to an extent different from how water interacts with it.[23,48–50] The interactions with cosolute represented in our model should therefore be regarded as site specific potentials of mean force. Further, cosolutes in the simulation were assumed to occupy one lattice site. The algorithm can be easily extended to also include cosolutes of various other sizes, by considering the change in chemical potential for solutions containing different sized cosolutes and solvent. However simple, the simulation still captures the essential effect of cosolute interacting with protein, namely the change in (free) energy upon association. The model serves as a practical demonstration of our technique, a powerful approach allowing us direct access to the effect of cosolute on protein stability.

While we model proteins in a very simple manner, certain fundamental features of cosolute denaturation and stabilization emerge. Experimentally, it is well established that many proteins undergo the chemical denaturation process with increased amounts of added cosolute in one step, folded to unfolded. This feature is often not very sensitive to the type of denaturant used. Treating the stabilization or destabilization as a two-step process (natured–denatured), it is often found that the change in protein stability (reported as the change in the standard free energy for the denaturation process) is almost linear with the concentration of added cosolute.[56–58] This linearity has been a topic of extensive experimental and theoretical study, and is often used as a way of determining (through extrapolation) the stability of the native protein in the absence of cosolute. In the simulation, Figure 4, we indeed find for a large range of parameters a linear dependence of $\delta\Delta G$ on cosolute concentration.

Based on the Gibbs adsorption isotherm, our simulation strategy also allows us to follow the change in stability with added cosolute through cosolute association with protein. The idea, first used with reference to proteins by Wyman[40] and Tanford[19] then Schellman[14,22,38] and others,[59,60] is related to that of Gibbs[31] for the free energy associated with creating an interface.[7] It states that the change in the energy required to create the interface with the addition of solute is in proportion to the excess of that solute in its vicinity. Later studies used this statement, together with different specific models, to extract from protein denaturation studies phenomenological parameters referred to as hydration, partition coefficients and the exposed surface area upon unfolding. In the simulation, we follow the relationship between protein stability and numbers of associated cosolutes. We demonstrate that at thermodynamic equilibrium, the linkage between cosolute adsorption and stability holds, and thus the extent of exposed protein surface with cosolute concentration can be followed (Fig. 6).

By preferentially interacting with the protein, the cosolutes change the ensemble probability of the different protein configurations, hence stabilizing or destabilizing different states. The stabilizing (or destabilizing) of the associating cosolute (the work of inserting protein into a cosolute- containing medium) is in proportion to the number of cosolutes associated with the protein. Further, in the language of Wyman linkage, cosolutes' *relative* stabilizing action of the native versus denatured states is in proportion to the *difference* in the number of cosolute molecules adsorbed (or depleted) in the two states [eq. (8)]. Independent of models and definitions of denatured, native, or any other state, by evaluating the difference in the number of excess cosolutes associated with different states at different cosolute concentrations, we are able to determine the change in their relative stability.

In contrast to simple adsorption, where the number of available surface sites is fixed, proteins are flexible and can respond to conditions set by solution by folding/unfolding to expose less or more sites for adsorption. We find that the effect of inclusion or exclusion of cosolute from the protein surface is to further increase or decrease the exposed part of the protein's chain so as to allow more or less surface to interact, as shown in Figure 5. For highly excluded cosolute, changes in protein stability are simply related to the change in the number of cosolute excluding (solvent) lattice sites upon denaturing. For cosolute interacting favorably with H/P, the protein may respond in one step - native to unfolded - or in a series of intermediate unfolding steps, depending on cosolute, as shown in Figure 5. The ensemble of denatured configurations found in simulation also differs in response to different cosolutes, reflecting the preferential interactions between each cosolute and the protein.

Experimentally, some proteins are found to undergo denaturation through one or more stable configurations (intermediates) in two or more denaturation steps.[1,61–63] The structure of the intermediates depends not only on the protein sequence, but also on the type of cosolute used; an example is the 'molten globule' state of some acid-denatured proteins.[64,65] In the past decade, important insight has been gained concerning the complexity of the denatured states.[66–69] Once thought to be random, it has now been shown that in many cases residual structure is maintained in the denatured state and that this structure is also dependent on the type of cosolute used.[70]

While stability of the simulated protein is directly reflected in numbers bound, ensuing configurational states and the structure of protein intermediates depend more sensitively on the specifics of interaction parameters. Similar denaturation intermediates result from exposure

to weakly and non-specifically interacting denaturing cosolutes as from thermal denaturation [compare Figs. 7 and 8(a)]. On the other hand, we find more complex behavior with strongly binding cosolutes, or ones that interact preferentially with a particular monomer type. Such cosolutes promote the formation of distinctly different intermediates from those found in thermal denaturation; compare the configuration in Figure 2(b) for cosolute interacting strongly only with P monomers to Figure 2(c and d) for weak binders to both P and H. The denatured state is thus shown to possess structure, which depends on the type of cosolute used.

Finally, we note that the simulation method can be extended to kinetic (e.g. molecular dynamic) studies. However, while for evaluation of thermodynamic properties this method is exact, relaxation times must be considered in kinetic studies. Only if cosolute–protein relaxation times are much faster than global protein rearrangement is the algorithm directly applicable.

ACKNOWLEDGMENTS

We thank Horia Petrache, Don Rau and Rudi Podgornik for helpful discussions and comments.

REFERENCES

1. Dill KA, Shortle D. Denatured states of proteins. Annu Rev Biochem 1991;60:795–825.
2. Alonso DOV, Dill KA. Solvent denaturation and stabilization of globular proteins. Biochemistry 1991;30:5974–5985.
3. Shortle D. The denatured state (the other half of the folding equation) and its role in protein stability. FASEB J. 1996;10:27–34.
4. Dill KA. Polymer principles and protein folding. Prot Sci 1999;8: 1166–1180.
5. Timasheff SN. Control of protein stability and reactions by weakly interacting cosolvents: the simplicity of the complicated. Adv Protein Chem 1998;51:355–432.
6. Baldwin RL. How Hofmeister ion interactions affect protein stability. Biophys J 1996;71:2056–2063.
7. Parsegian VA. Protein-water interactions. Int Rev Cyt 2002;215:1–31.
8. Minton AP. Effect of a concentrated "inert" macromolecular cosolute on the stability of a globular protein with respect to denaturation by heat and by chaotropes: A statistical-thermodynamical model. Biophys J 2000;78(1):101–109.
9. Xie GF, Timasheff SN. Mechanism of the stabilization of ribonuclease a by sorbitol: Preferential hydration is greater for the denatured than for the native protein. Prot Sci 1997;6(1):211–221.
10. Bolen DW, Baskakov IV. The osmophobic effect: Natural selection of a thermodynamic force in protein folding. J Mol Biol 2001;310: 955–963.
11. Courtenay ES, Capp MW, Anderson CF, Record MT. Vapor pressure osmometry studies of osmolyte-protein interactions: Implications for the action of osmoprotectants in vivo and for the interpretation of "osmotic stress" experiments in vitro. Biochemistry 2000;39:4455–4471.
12. Courtenay ES, Capp MW, Record MT. Thermodynamics of interactions of urea and guanidinium salts with protein surface: Relationship between solute effects on protein processes and changes in water-accessible surface area. Prot Sci 2001;10:2485–2497.
13. Schellman JA. Protein stability in mixed solvents: A balance of contact interaction and excluded volume. Biophys J 2003;85:108–125.
14. Schellman JA. The thermodynamic stability of proteins. Ann Rev Biophys Biophys Chem 1987;16:115–137.
15. Pace CN, Laurents DV, Thomson JA. pH dependence of the urea and guanidine hydrochloride denaturation of ribonuclease A and ribonuclease T1. Biochemistry 1990;29:2564–2572.
16. Arakawa T, Timasheff SN. Stabilization of protein structure by sugars. Biochemistry 1982;21:6526–6544.
17. Qu Y, Bolen CL, Bolen DW. Osmolyte-driven contraction of a random coil protein. Proc Natl Acad Sci USA 1998;95:9268–9273.
18. Tanford C. Isothermal unfolding of globular proteins in aqueous urea solutions. J Am Chem Soc 1964;86:2050–2059.
19. Tanford C. Protein denaturation part c: Theoretical models for the mechanism of denaturation. Adv Protein Chem 1970;24:1–95.
20. Wu H. Studies on denaturation of proteins xiii. a theory of denaturation. Chin J Physiol 1931;5:321–344.
21. Yang M, Ferreon ACM, Bolen DW. Structural thermodynamics of a random coil protein in guanidine hydrochloride. Prot Struc Fun Gen 2000;(Suppl. 4):44–49.
22. Schellman JA. Macromolecular binding. Biopolymers 1975;14:999–1018.
23. Dwyer DS. Molecular simulation of the effects of alcohols on peptide structure. Biopolymers 1998;49:635–645.
24. Frenkel D, Smit B. Understanding molecular simulations: from algorithms to applications. San Diego: Academic Press, 1996.
25. Caflisch A, Karplus M. Acid and thermal denaturation of barnase investigated by molecular dynamics simulations. J Mol Biol 1995;252:672–708.
26. Colonna-Cesari F, Sander C. Excluded volume approximation to protein-solvent interaction. The solvent contact model. Biophys J 1990;57:1103–1107.
27. Wesson L, Eisenberg D. Atomic solvation parameters applied to molecular dynamics of proteins in solution. Protein Sci 1992;1:227–235.
28. Fraternali F, van Gunsteren WF. An efficient mean solvation force model for use in molecular dynamics simulations of proteins in aqueous solution. J Mol Biol 1996;256:939–948.
29. Lazaridis T, Karplus M. Effective energy function for proteins in solution. Prot Struc Func Gen 1999;35:133–152.
30. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AN, Teller E. Equation of state calculations by fast computing machines. J Chem Phys 1953;21:1087–1092.
31. Gibbs JW. On the equilibrium of heterogeneous substances. In: Bumstead HA, van Name RG, editors. The scientific papers of J. Willard Gibbs. Ox Bow: Woodbridge, 1993. Vol. 1, sect. 3, p. 55–353.
32. Baptista A, Martel P, Petersen S. Simulation of protein conformational freedom as a function of pH: constant-pH molecular dynamics using implicit titration. Proteins 1997;27:523–544.
33. Walczak AM, Antosiewicz J. Langevin dynamics of proteins at constant pH. Phys Rev E 2002;66:051911–1–8.
34. Borjesson U, Hunenberger P. explicit-solvent molecular dynamics simulation at constant pH: Methodology and application to small amines. J Phys Chem 2001;114:9706–9719.
35. Burgi R, Kollman P, van Gunstern W. Simulation of proteins at constant pH: An approach combining molecular dynamics and monte carlo simulation. Proteins 2002;47:469–480.
36. Lau KF, Dill KA. A lattice statistical-mechanics model of the conformational and sequence-spaces of proteins. MACROMOLECULES 1989;22:3986–3997.
37. Miller DW, Dill KA. Ligand binding to proteins: The binding landscape model. Prot Sci 1997;6:2166–2179.
38. Schellman JA. Solvent denaturation. Biopolymers 1978;17:1305–1322.
39. Hill TL. Introduction to statistical thermodynamics. New York: Addison-Wesley, 1960.
40. Wyman J. Adv Protein Chem 1964;19:223–286.
41. Micheletti C, Banavar JR, Maritan A. Conformations of proteins in equilibrium. Phys Rev lett 2001;87:088102.
42. Gupta P, Hall CK. Effect of solvent conditions upon refolding pathways and intermediates for a simple lattice protein. Biopolymers 1997;42(4):399–409.
43. Gupta P, Hall CK, Voegler AC. Effect of denaturant and protein concentrations upon protein refolding and aggregation: A simple lattice model. Prot Sci 1998;7(12):2642–2652.
44. Nguyen HD, Hall CK. effect of rate of chemical or thermal renaturation on refolding and aggregation of a simple lattice protein. Biotec Bioeng 2002;80(7):823–834.
45. Flory PJ. Statistical mechanics of chain molecules. New York: Wiley-Interscience, 1969.
46. de Gennes PG. Scaling concepts in polymer physics. Ithaca: Cornell University, 1979.
47. Grosberg AY, Khokhlov AR. Statistical physics of macromolecules. New York: AIP Press, 1994.
48. Tobi D, Elber R, Thirumalai D. The dominant interaction between

peptide and urea is electrostatic in nature: A molecular dynamics simulation study. Biopolymers 2003;68:359–369.
49. Wallqvist A, Covell DG, Thirumalai D. Hydrophobic interactions in aqueous urea solutions with implications for the mechanism of protein denaturation. J Am Chem Soc 1998;120:427–428.
50. Bennion BJ, Daggett V. The molecular basis for the chemical denaturation of proteins by urea. Proc Natl Acad Sci 2003;100: 5142–5147.
51. Sali A, Shakhnovich E, Karplus M. How does a protein fold? Nature 1994;369:248–251.
52. Dinner A, Sali A, Karplus M, Shakhnovich E. Phase-diagram of a model protein derived by exhaustive enumeration of the conformations. J Chem Phys 1994;101:1444–1451.
53. Klimov DK, Thirumalai D. Lattice models for proteins reveal multiple folding nuclei for nucleation-collapse mechanism. J Mol Biol 1998;282:471–492.
54. Combe N, Frenkel D. Phase behavior of a lattice protein model. J Chem Phys 2003;118(19):9015–9022.
55. Pokarowski P, Kolinski A, Skolnick J. A minimal physically realistic protein-like lattice model: Designing an energy landscape that ensures all-or-none folding to a unique native state. Biophys J 2003;84:1518–1526.
56. Aune K, Tanford C. Thermodynamics of the denaturation of lysozyme by guanidine hydrochloride ii. dependence on denaturant concentration at 25°. Biochemistry 1969;8:4586–4590.
57. Myers JK, Pace CN, Scholtz JM. Denaturant in values and heat capacity changes: relation to changes in accessible surface area of protein unfolding. Prot Sci 1995;4:2138–2148.
58. Bolen DW, Santoro MM. Unfolding free energy changes determined by the linear extrapolation method. 2. Incorporation of $\Delta G°_{n-u}$ values in a thermodynamic cycle. Biochemistry 1988;27: 8069–8074.
59. Baskakov IV, Bolen DW. Monitoring the sizes of denatured ensembles of staphylococcal nuclease proteins: Implications regarding in values, intermediates, and thermodynamics. Biochemistry 1998;37:18010–18017.
60. Kita Y, Arakawa T, Y, LT, Timasheff SN. Contribution of the surface free-energy perturbation to protein solvent interactions. Biochemistry 1994;33:15178–15189.
61. Dabora J, Marqusee S. Equilibrium unfolding of Escherichia coli ribonuclease H: Characterization of a partially folded state. Prot Sci 1994;1401–1408.
62. Barrick D, Baldwin RL. Three-state analysis of sperm whale apomyoglobin folding. Biochemistry 1993;27:3790–3796.
63. Pande VS, Grosberg AY, Tanaka T, Rokhsar DS. Pathways for protein folding: Is a new view needed? Curr Op Struc Biol 1998;8:68–79.
64. Aune KC, Salahuddin A, Zarlengo MH, Tanford C. Evidence for residual structure in acid and heat denatured proteins. J Biol Chem 1967;242:4486–4489.
65. Carra JH, Anderson EA, Privalov PL. Thermodynamics of staphylococcal nuclease denaturation. I. The acid denatured state. Prot Sci 1994;3:944–951.
66. Shortle D, Ackerman MS. Persistence of native-like topology in a denatured protein in 8 M urea. Science 2001;293:487–489.
67. Doniach S. Changes in biomolecular conformation seen by small angle x-ray scattering. Chem Rev 2001;101:1763–1778.
68. Mayor U, Grossmann JG, Foster NW, Freund SMV, Fersht AR. The denatured state of engrailed homeodomain under denaturing and native conditions. J Mol Biol 2003;333:977–991.
69. Zagrovic B, Snow CD, Khaliq S, Shirts MR, Pande VS. Native-like mean structure in the unfolded ensemble of small proteins. J Mol Biol 2002;323:153–164.
70. Ohnishi S, Shortle D. Effects of denaturants and substitutions of hydrophobic residues on backbone dynamics of denatured staphylococcal nuclease. Prot Sci 2003;12:1530–1537.